

Full Report

HeppSY Cohort Evaluation

Contact Hours and Expectations of
Applying to HE Aged 18 or 19

October 2020

Authors: Daniel Fletcher and Greg Brown



Contents

3 Introduction and Context

- 3 Local Impact Evaluation
- 3 Aims and Objectives
- 4 Hypothesis and Research Question

5 Methodology

- 5 Data
- 6 Matching Process
- 6 Measures
- 7 Sample
- 8 Analytical Method

9 Analysis and Results

- 9 Descriptive Statistics
- 12 Regression Analysis
- 15 Exploratory Analysis

17 Discussion

- 18 Limitations

19 Conclusions and Recommendations

- 19 Headline Conclusions
- 19 Recommendations for HeppSY Practice
- 20 Recommendations for Future Research and Evaluation

21 References

22 Appendix 1

Introduction and Context

The Higher Education Progression Partnership South Yorkshire (HeppSY) is part of the national Uni Connect programme (UCP), funded by the Office for Students (OfS), to help school and college students aged 13-19 across South Yorkshire who are most at risk of missing out on higher education (HE). HeppSY is working in partnership with Sheffield Hallam University, The University of Sheffield and South Yorkshire schools and colleges.

The Uni Connect programme aims to increase the HE participation rate of students from underrepresented backgrounds, and to support young people in making well-informed decisions about their future (OfS, 2020). A key part of the programme is also to evaluate the effectiveness of outreach provision in achieving these goals.

Local Impact Evaluation

The local impact evaluation at HeppSY is split into two overarching strands: cohort-level analyses, and programme case studies. The purpose of cohort-level analyses are to investigate how effective the programme has been overall in achieving its aims, by determining associations and causal relationships between participation in outreach activity and outcomes related to HE. This type of analysis is important as it incorporates a broad range of the outreach activity that is delivered as part of the programme and is well suited to tracking outcomes over a long period of time for large sample sizes. The limitation is that the diversity of the activity profiles of different students means that even if the data suggest the overall programme was effective, it is difficult to determine which features were necessary and sufficient in achieving a given outcome.

To address this concern, a series of programme level impact evaluations will be undertaken. These will be small-scale, well-controlled pieces of evaluation on specific, sustained programmes of activity. The aim is to provide robust evidence for the effectiveness of specific interventions, which are based on a detailed theory of change model. This is further supported by ongoing evaluation of the partnership's core offer, through case-study evaluations of purposively sampled activities across the four strands of the programme.

This evidence will offer the programme and our funders an insight into what type of interventions have the greatest impact for different target student groups.

Aims and Objectives

The focus of the present analysis is on the cohort-level strand of the impact evaluation. Ultimately, the effectiveness of the programme at a total cohort level is likely to be best assessed through objective HE participation outcomes, using data from the Higher Education Statistics Agency (HESA). However, due to the timeframes involved in this data becoming available, outcomes for the Year 9-11 students that participated in activity during the first full academic year of the programme (2017-18) will not be available before the programme is due to end in July 2021. Therefore, there is a clear need to utilise measures that are predictive of future HE entry, and which are theoretically expected to be influenced by outreach activity.

Drawing on data from Wave 1 and Wave 2 of the CFE annual learner survey, and HeppSY and partner activity data held in the Higher Education Access Tracker (HEAT); the aim of this project was to investigate the association between the number of contact hours participated in and changes in self-reported likelihood of applying to HE, for students that were in Year 10 during the first wave of the survey (Year 11 at Wave 2).

The evidence that expectation of applying to HE is a good predictor of actual applications is somewhat mixed. Combining all likely to apply responses into a binary category does appear to overestimate what actual entry rates are likely to be (HeppSY, 2020), however longitudinal analysis by (Anders & Micklewright, 2015) suggests that by age 16, a strong expectation of progressing to HE is a very good predictor of actual progression. Measuring expectation of applying to HE on an ordered scale, which allows strength of intention to be expressed, was therefore considered to be a suitable indicator for evaluation.

Hypothesis and Research Question

Cross-sectional comparisons using CFE survey data have shown that Year 11 students have lower expectations of applying to HE than either Year 9 or Year 10 students (HeppSY, 2020), and this relationship has also been observed longitudinally in the wider literature (Anders & Micklewright, 2015). The present analysis was therefore an opportunity to assess whether level of participation in the HeppSY programme was associated with a reduced likelihood of a negative shift in expectations between Year 10 and Year 11.

As such, the primary research question was:

After controlling for prior expectations, what is the association between the number of HeppSY contact hours received by students in Year 10, and their expectation of applying to HE in Year 11?

This analysis will be supported by further work utilising the same dataset, but focussing on different questions from the learner survey, to give more detailed insight into the impact of specific HeppSY interventions across the four strands of the programme (HE knowledge, career knowledge, confidence and resilience, attainment).

Methodology

Data

CFE Survey

The student survey data used here was collected as part of the Wave 1 (W1) and Wave 2 (W2) CFE Annual Learner Survey. W1 responses were gathered from students in HeppSY partner schools and colleges in Autumn 2018, and W2 data was collected in Winter 2019. HeppSY co-ordinated with Key Points of Contact (KPOC) in partner institutions, who administered the survey to their students on our behalf. The W1 and W2 survey datasets contained data for 10786 and 12246 students, respectively.

HEAT

The HEAT dataset was based on HeppSY and partner (UoS, SHU, Hepp) HEAT exports of student records and OfS funded activity data for Phase 1. Student records were matched across the four export files using a concatenation of the first three letters of student first name, full surname, date of birth, and postcode. The HEAT dataset included information on the cumulative number of contact hours, total activities, and the count of activity by type that each student had participated in.

School/College Registers

Each academic year, partner institutions provide HeppSY with a register dataset including all students in Year 9-13, and their assigned HE potential coding (see the Measures section below for a description). The 2018-19 register was used in the present analysis to link HE potential coding to students in the matched survey dataset.

Matching Process

The process used for matching students across datasets is described below. Note that the present analysis focused on students who were in Year 10 at W1, and therefore only used a subset of the overall matched dataset.

Individual students were matched across the two waves of survey data using fuzzy matching in Excel Power Query. Matches of 90% or higher were accepted, subject to random spot-checks of the data. This resulted in 1892 matches. Student records were merged so that every case had a W1 and a W2 response for each survey question.

The linked survey responses were then matched to individual student activity records in the HEAT dataset. Activity was filtered by date, so that only engagements that took place in-between the close of the W1 survey and the opening of W2 were included in the matching process. Fuzzy matching of the linked survey and HEAT datasets resulted in 1496 matches.

The 396 linked survey records that could not be matched to the HEAT dataset were either due to data quality issues, or because a student had not participated in any activity between survey waves. As non-matches do not necessarily indicate non-participation, these records were excluded from the dataset rather than being included as zero activity students.

Finally, the 1496 students that had been matched to both survey waves and to the HEAT dataset were matched to HeppSY partner institution register data, to add the HE potential coding that schools/colleges had assigned to students. This resulted in 1387 matches, which formed the final dataset of students.

Measures

Likelihood of applying to HE (Survey W1 and W2)

Students were asked, 'how likely are you to apply to higher education at age 18 or 19?' Responses were given on a six-point ordered scale (1 – definitely won't apply, 2 – very unlikely, 3 – fair unlikely, 4 – fairly likely, 5 – very likely, 6 – definitely will apply), with a don't know response option also available.

Motivation (Survey W1):

Students were asked to rate their level of agreement with the statement, 'I am motivated to do well in my studies.' Responses were given on a 5-point Likert scale (1 – strongly disagree, 2 - slightly disagree, 3 – neither agree nor disagree, 4 – slightly agree, 5 – strongly agree).

Contact hours (HEAT)

Reflects the cumulative number of hours of HeppSY outreach a student had participated in between the two survey waves.

HE Potential Coding (School Registers)

HE potential coding is assigned to students by KPOCs at HeppSY institutions, based on whether a student lives in a target postcode, and whether they believe that the student is academically capable of progressing to HE.

Green = student is from a HeppSY target postcode and has the potential to progress to HE.

Red = student is from a HeppSY target postcode but does not have the potential to progress to HE.

Non-UCP = student is not from a HeppSY target postcode.

Sample

There were 679 Year 10 (W1) students in the dataset. The following exclusion criteria were applied: 1) don't know responses to the likelihood of applying to HE question at W2 or missing responses at either wave (N = 111); 2) missing response for gender at both waves (N = 9); 3) contact hours over 25 hours (~ 2 S.D above the mean, N = 30); 4) students from schools with less than five respondents¹ (N = 19).

The final sample included 514 students. See Table 1 for a breakdown of respondents by gender and HE potential. Students from 20 different schools were included in the sample (school N after exclusions ranged from 6 - 62 students).

Table 1. Sample size by gender and HE potential coding.

Gender	HE Potential			Total
	Green	Red	Non-UCP	
Male	93	30	82	205
Female	159	29	106	294
Prefer not to say / Other	9	1	5	15
Total	264	60	199	514

.....

1 Students from schools with fewer than five respondents were excluded to avoid quasi-separation in the logistic regression model, which occurred where all students within a school provided the same response on the outcome variable.

Analytical Method

As the outcome variable was on an ordinal scale (points on the scale are in order, but the interval between each point is not assumed to be equal), ordinal logistic regression was used. Ordinal logistic regression is an extension of binary logistic regression, which models the likelihood of a specified outcome occurring from a dichotomous variable, given a set of independent predictor variables (O'Connell, 2006).

In a binary logistic regression model, the effect of each independent predictor is often expressed in terms of odds ratios (ORs). For categorical variables, ORs represent the comparative likelihood that participants in a particular group of an independent predictor (e.g. male respondents) will have the target outcome, compared to participants in a specified reference category (e.g. female respondents). An OR of one indicates that participants from both groups are equally likely to have the target outcome, an OR of less than one indicates that participants in a particular group are less likely to have the target outcome compared to the reference group, and an OR of greater than one indicates that participants in a particular group are more likely to have the target outcome.

For continuous predictor variables, ORs reflect the change in the odds of participants having the target outcome for each additional unit of the predictor (e.g. each additional contact hour).

Ordinal logistic regression uses these principles but is extended to reflect an outcome variable with more than two response options, and with a logical order to the scale. The type of ordinal regression used here (with the assumption of proportional odds) can be conceptualised as a series of binary logistic regressions across cumulative, dichotomous partitions of the data (O'Connell, 2006). For example, on the likelihood of applying to HE scale used in the present analysis, the following five partitions of scores exist: 1) very unlikely and higher 2) fairly unlikely and higher, 3) fairly likely and higher, 4) very likely and higher, and 5) definitely will apply; with the alternative outcome in the dichotomy for each partition being scores below the specified cut-off. A common OR for each predictor is given to explain the relationship with the outcome variable across all cumulative partitions.

Analysis and Results

Descriptive Statistics

HEAT Activity Data

The median contact hours for the sample was 3.83 hours, ranging from 0.25 – 23.66 hours. Median contact hours by gender and HE potential are displayed in Table 2. Females had nearly twice as many contact hours as males, Green students had more contact hours than Red students, and both Green and Red students had over three times the amount of contact hours as Non-UCP students.

Table 2. Median contact hours by gender and HE potential coding.

Male	Female	Green	Red	Non-UCP
2.83	5.33	7.38	6.25	2.00

Details of participation by activity type are displayed in Table 3. General HE information was the most frequently participated in activity, with 89% of students engaging in this activity type at least once, and a median of two engagements amongst these students. All students participated in at least one activity, with a median of four engagements per student across all activities.

Table 3. Student engagement by activity type.

Activity Type	Participated in Activity Type (N =)		Median*
	Yes	No	
General HE Information	458	55	2
Campus Visit	109	405	1
HE Subject Insight	78	436	1
Mentoring	148	366	1
Study Skills and Attainment	192	322	1
Summer School	6	508	1
Other Activity Type	240	274	1
All Activities	514	0	4

*Refers to the median number of engagements for students that participated in that activity type, and therefore does not include zero engagement students.

Survey Data

At W1, a high proportion of students expressed an intention to apply to HE, with 72.2% responding that they were fairly likely, very likely, or definitely would apply. A lower proportion of students expressed an intention to apply at W2, with 66.2% indicating that they were likely to apply (see Fig. 1).

Figure 1. Likelihood of applying to HE – Wave 1 and Wave 2 Responses.

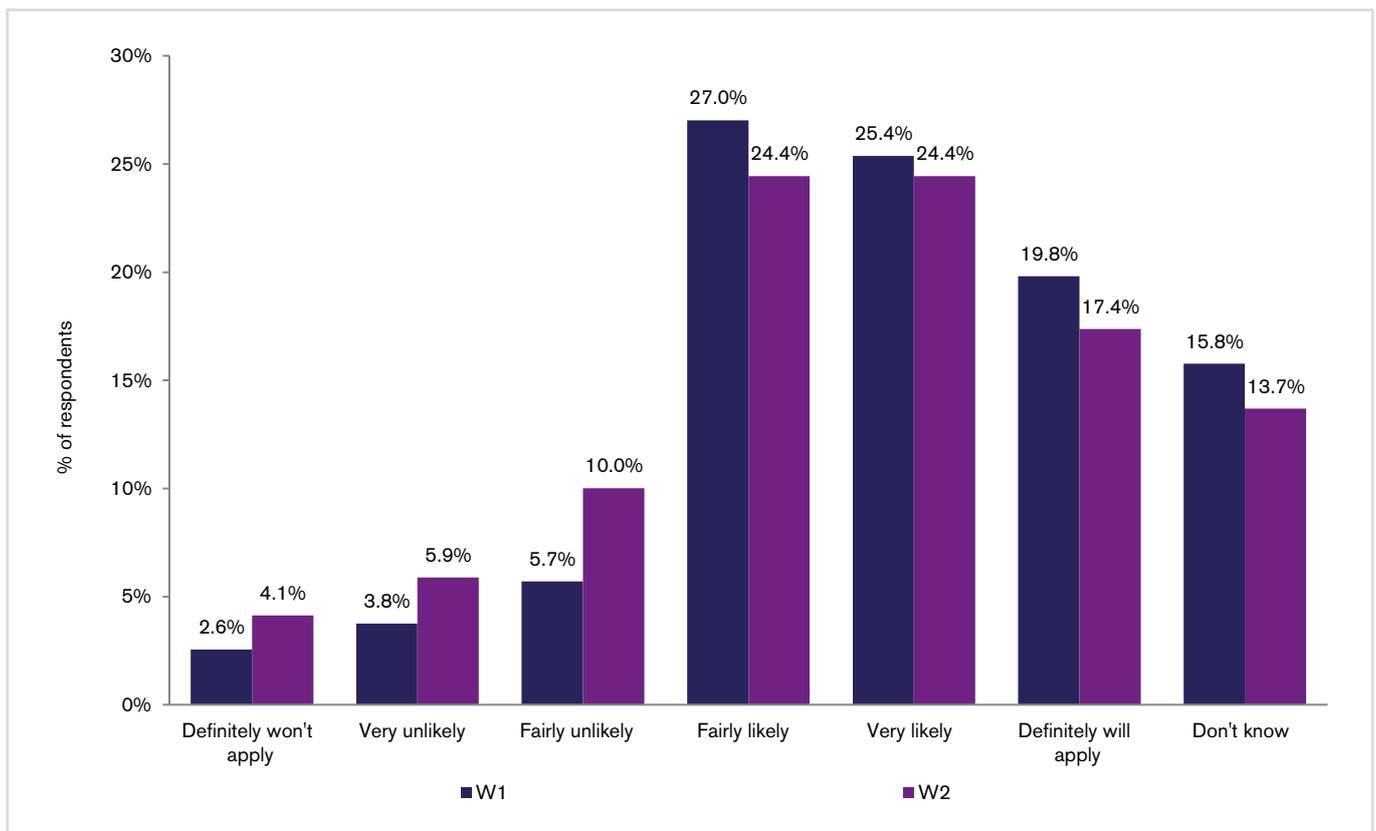
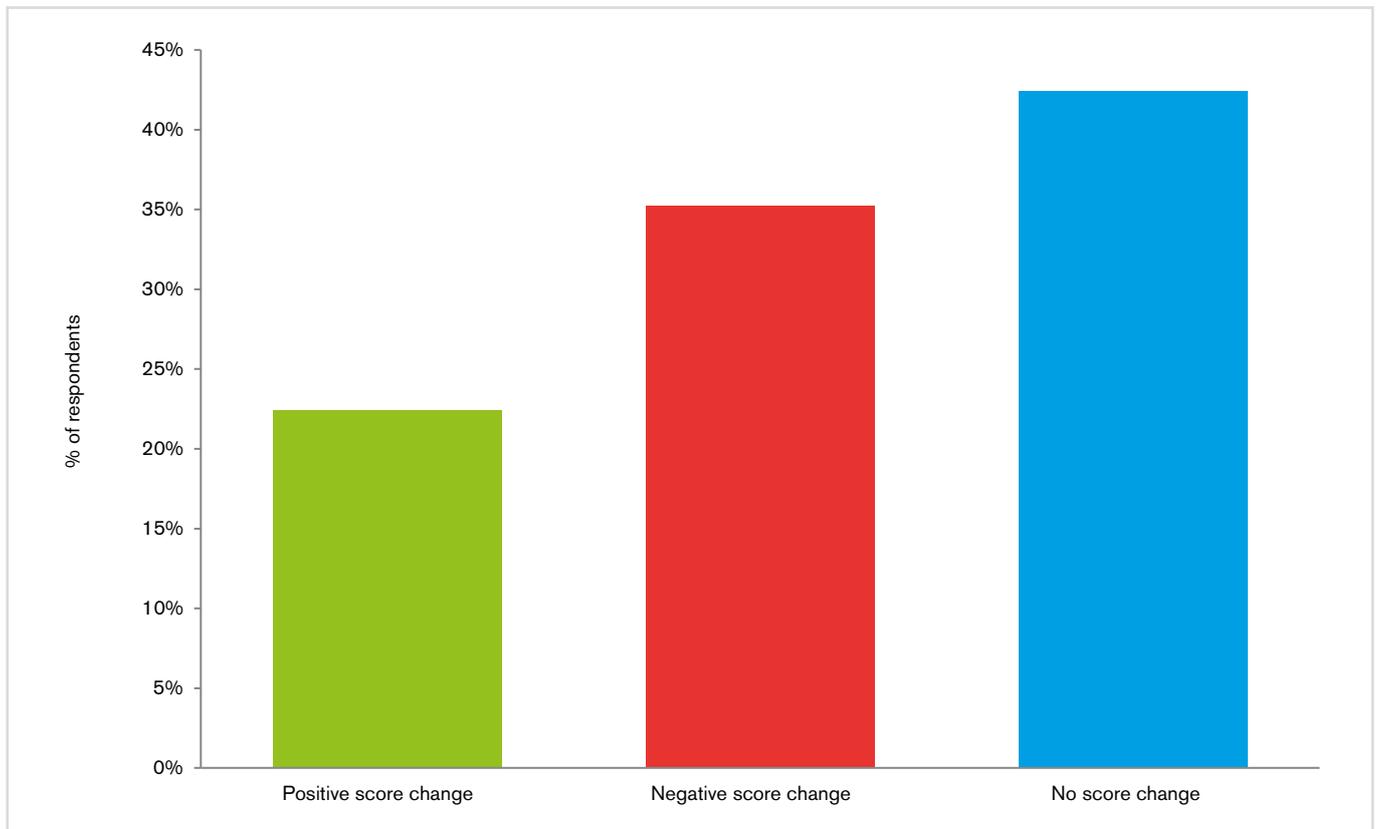


Figure 2 shows the proportion of students that provided a higher W2 response compared to W1 on the six-point ordered scale, the proportion that gave a lower score, and the proportion that gave the same response across both waves. Students were around 1.6 times more likely to indicate a reduction in expectations of applying to HE between waves compared to an increase in expectations, though no change in response was the modal outcome. This downward shift in expectations observed here for matched student responses is in line with the cross-sectional data for the total survey sample (HeppSY, 2020).

Figure 2. Change in likelihood of applying to HE between waves.



Note. N = 460. Excludes students that gave a don't know response at either wave.

Regression Analysis

A summary of the regression models is displayed in Table 4, and the odds ratios (ORs) and significance values for predictor variables are shown in Table 5 (log odds and standard errors are in Appendix 1).

Model 1

In the first iteration of the model, HE potential coding, gender, W1 likelihood of applying to HE, and contact hours were entered as predictor variables. Both Green (OR = 1.71; 95% CI [1.01, 2.88]) and Non-UCP (OR = 2.29; 95% CI [1.32, 3.98]) students were significantly more likely than Red students to have higher W2 likelihood of applying to HE responses. Males were less likely than females to have higher W2 responses, OR = 0.64, 95% CI [0.46, 0.90]. There was a positive, but non-significant association between contact hours and W2 responses, OR = 1.02, 95% CI [0.99, 1.05].

Model 2

In the second model, school was added as a predictor. The purpose of adding school was not to directly interpret the effect of each school on the outcome variable, but instead to interpret the effect of other predictor variables, having controlled for school-level effects. In this model, there was a significant relationship between contact hours and W2 responses, with each additional contact hour associated with 1.05 times greater odds of a student having a higher W2 likelihood of applying to HE response, OR = 1.05, 95% CI [1.01, 1.10]. Both Green (OR = 1.49, 95% CI [0.83, 2.68]) and Non-UCP (OR = 1.76, 95% CI [0.90, 3.45]) students were still associated with greater W2 likelihood of applying relative to Red students, however the relationship was no longer significant. Gender remained a significant predictor, with males less likely than females to have higher W2 responses, OR = 0.60, 95% CI [0.42, 0.85].

Model 3

In the third iteration, student's self-reported motivation to do well in their studies at W1 was added to the model. Note that not all students answered this question, so including it as a predictor slightly reduced the sample size. Motivation was not significantly associated with W2 responses (OR = 1.12, 95% CI [0.91, 1.38]), and its inclusion in the model did not substantially change the parameter estimates or significance of other variables.

Table 4. Model summaries.

	N	Pseudo R-square*	Chi-square	df	p-value
Model 1	514	.33	198.13	12	<.01
Model 2	514	.39	242.84	31	<.01
Model 3	508	.39	235.13	32	<.01

*Nagelkerke

df = model degrees of freedom

Table 5. Model parameter estimates.

Variable	Model 1		Model 2 (inc school effects)		Model 3 (inc school effects and motivation)	
	OR	p-value	ORs	p-value	OR	p-value
Contact Hours	1.02	.30	1.05	.02	1.05	.03
HE Potential: Green ^a	1.71	.04	1.49	.19	1.50	.18
HE Potential: Non-UCP ^a	2.29	<.01	1.76	.10	1.70	.13
Male ^b	0.64	.01	0.60	<.01	0.60	<.01
W1 – Definitely won't apply ^c	0.32	.05	0.25	.02	0.27	.03
W1 - Very unlikely ^c	0.38	.04	0.31	.02	0.32	.02
W1 - Fairly unlikely ^c	0.96	.93	1.16	.72	1.34	.50
W1 - Fairly likely ^c	2.02	.01	2.17	<.01	2.10	.01
W1 - Very likely ^c	5.05	<.01	5.21	<.01	4.94	<.01
W1 - Definitely will apply ^c	19.33	<.01	23.73	<.01	21.88	<.01
Motivation (W1)					1.12	.27

^a Reference category - HE Potential: Red

^b Reference category – Female

^c Reference category – 'Don't know' W1 responses

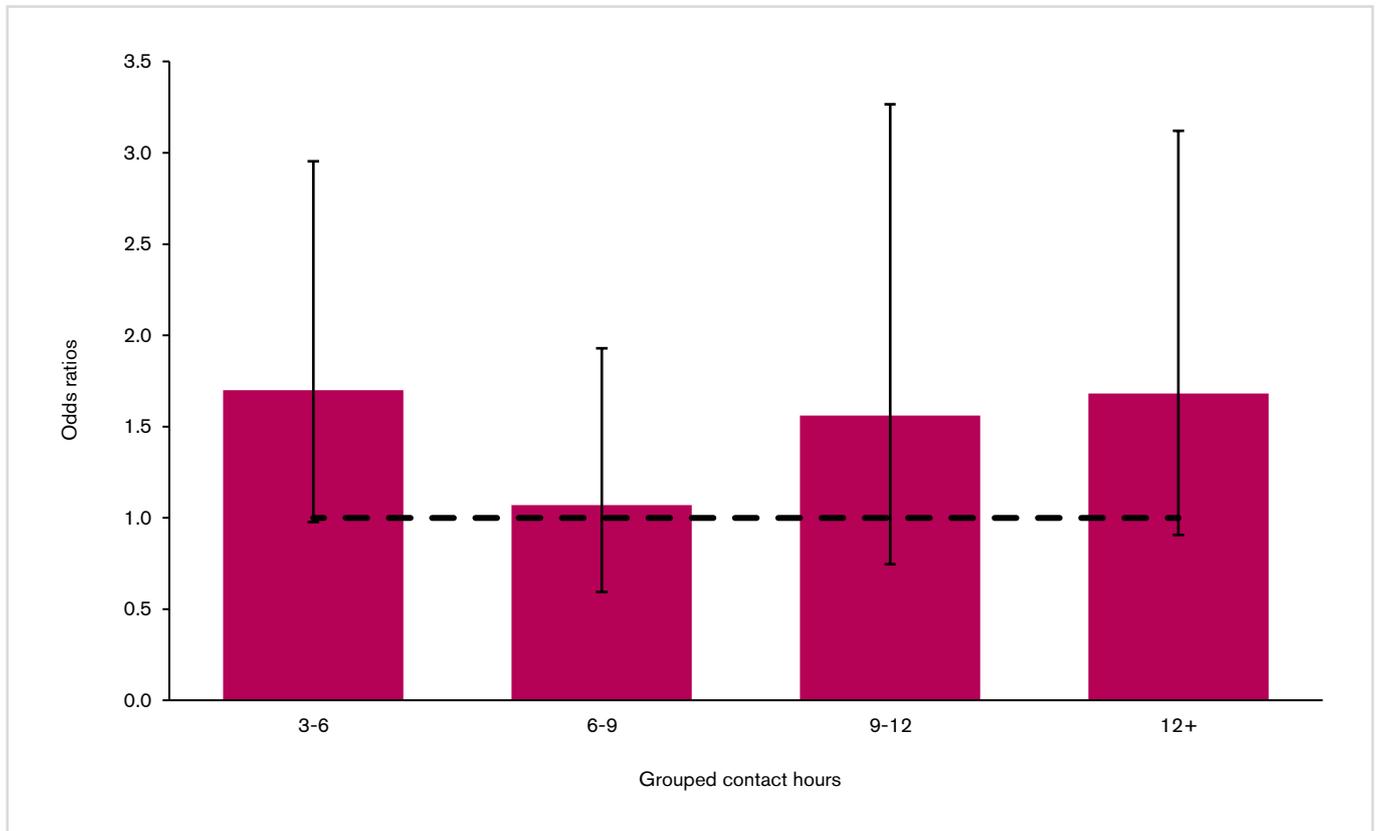
Exploratory Analysis

The association between contact hours and W2 likelihood of applying to HE that was demonstrated in Model 2 and 3 is only open to a linear interpretation, with each additional contact hour ostensibly producing the same incremental effect on the odds of a student having a higher W2 response. For example, the model predicts that an additional contact hour for students with three hours, and an additional contact hour for students with 15 hours would both increase the odds of a higher W2 response by 1.05 times.

However, it is possible that the effect of an additional contact hour changes depending on the number of hours already received. There may also be a range of contact hours that seem to be particularly effective (or ineffective). To investigate this, additional analysis was conducted using contact hour groupings. A grouped contact hours variable was created, categorising hours into five groups: 0-3 hours (N = 235), 3.01-6 hours (N = 66), 6.01-9 hours (N = 82), 9.01-12 hours (N = 37), 12.01+ hours (N = 94). These categories were chosen on the basis that they provided equal ranges of contact hours, and acceptable sample sizes.

An ordinal regression was conducted on W2 likelihood of applying to HE, with W1 likelihood of applying to HE, gender, HE potential coding, school, and grouped contact hours (0-3 hours as the reference category) added as predictor variables. Self-reported motivation was not included, as the regression analyses outlined in the section above indicated that it does not add to the explanatory power of the model (see pseudo R-square in Table 4). The odds ratios are displayed in Figure 3.

Figure 3. Odds ratios by contact hours group.



Reference category = 0-3 contact hours

Error bars reflect 95% confidence intervals

Dashed line reflects even odds

Students with 3-6 contact hours were 1.7 times more likely to have a higher W2 response relative to the reference group (less than three hours). Higher contact hour categories (6-9, 9-12, 12+) were not associated with greater relative ORs than the 3-6-hour category.

Discussion

The aim of the present analysis was to explore the association between contact hours and self-reported likelihood of applying to HE at Wave 2 of the CFE annual learner survey, after controlling for responses at Wave 1. A moderate, positive association was observed, but only when school effects were added to the model. The relationship remained after controlling for self-reported motivation, which did not have any substantial impact on the model.

Prior to the regression modelling, analysis of the descriptive statistics found that in line with cross-sectional analysis using the learner survey data (HeppSY, 2020), expectations of applying to HE reduced between waves for this matched sample of students. This supports the idea that a successful outcome of sustained outreach programmes may be to maintain already-high expectations of HE progression (particularly amongst pre-16 students), rather than requiring that an increase in expectations be observed.

Exploratory analysis was also conducted to determine whether a certain range of contact hours appeared to be particularly effective. Students with 3-6 contact hours were 1.7 times more likely to have a higher W2 response than students with less than 3 hours. Being in a contact hour group higher than 3-6 hours did not appear to increase the odds of a higher W2 response, particularly those receiving between 6-9 hours of contact. 3-6 hours of contact therefore appears to be the most efficient level of engagement, with the addition of incremental hours beyond this appearing to add little value with reference to our outcome variable of interest.

The finding that there was only an association between contact hours and W2 responses after school was added as a factor may have been caused by the clustering of schools at certain points along the contact hours scale. For example, a school that accounted for a large proportion of cases with high (low) contact hours, but whose students disproportionately had low (high) W2 response scores due to unobserved school-level influences, could have masked the association between contact hours and W2 responses that existed at a student level.

Interpretation of school-level effects becomes more complicated when you consider that the profile of outreach activity (target outcomes, intensity, specific sessions, etc.,) received is likely to vary far more between students from different schools compared to students within the same school. Therefore, what has been referred to as 'school-level effects' in the current paper is more accurately a combination of unobserved school-level variables and the profile of outreach activity, which cannot be disentangled with the available data.

Limitations

An important limitation of this analysis is that the sample was not random. Schools more closely engaged with HeppSY may have been more likely to facilitate student participation in the survey across both waves, and details on how schools select students to complete the survey are not always known. Ideally, all students would be given the opportunity to respond, but it is possible that schools selected specific classes to participate based on convenience or other factors.

Similarly, student response bias is a potential issue. Students who completed the survey across both waves may have differed in important ways from those that had the opportunity to participate but chose not to. For example, perhaps students who felt that HeppSY outreach had helped them to reconsider their HE options were more inclined to complete Wave 2 of the survey than students who had participated in HeppSY outreach but did not find it to be valuable. Issues with sample selection and response bias are highlighted by the disproportionate number of girls in the sample relative to boys.

The issue with using inferential statistics such as p-values and confidence intervals on data drawn from a non-random sample (Gorard, 2014) is also recognised. This information was included here, along with effect size, to give an indication of what effects appeared to be statistically robust. Full details of the statistics were provided to allow readers to make their own informed interpretations regarding the strength of the evidence.

An additional limitation is that attainment data was not included in the model. Attainment data on individual students from the National Pupil Database (NPD) is not currently available through HEAT for UCP consortia. While the HE potential coding given to students by schools was included, this is not a precise or objective measure, with different schools likely to use varying criteria in establishing the coding. Not including an objective measure of prior educational achievement such as KS2 or KS3 attainment is an important confound that precludes any causal claims being made from the data. Longitudinal analyses in the wider literature have found that attainment explains most of the variance in HE participation (Chowdry, Crawford, Dearden, Goodman, & Vignoles, 2014), and it is possible that attainment correlates with both contact hours and W2 likelihood of applying to HE, even after controlling for W1 responses (i.e., students with high prior attainment may have been more likely to maintain positive expectations around HE progression between waves compared to students with low prior attainment).

Conclusions and Recommendations

Ultimately, the HeppSY programme seeks to increase rates of HE progression amongst learners in designated target wards. As highlighted by Anders and Micklewright (2015), expectation of HE progression at age 16 is a strong predictor of actual progression. As part of the local impact evaluation, this research then sought to determine whether higher levels of engagement with the HeppSY programme between Year 10 and 11 (using contact hours as an indicator) maintained or increased student expectation of progression to HE at age 18 or 19.

This final section summarises the analysis undertaken and makes some recommendations for HeppSY practice moving forward, as well as for future research to unpick those issues uncovered.

Headline Conclusions

For students in Year 10 during the first wave of the survey, we found a moderate, positive association between contact hours and Wave 2 expectations of applying to HE, after controlling for Wave 1 responses (amongst other predictors). This association was statistically significant when school and motivation were included as further predictors.

This suggests that the more a Year 10 student interacts with the HeppSY programme, the greater the odds are that they will maintain or increase their expectations of progression to HE study at age 18 or 19 – a significant focus of the programme as whole.

Further, the exploratory analysis undertaken sought to identify a 'saturation point' at which further interaction with the programme offered little or no benefit against the outcome of interest. The exploratory model highlighted that students with 3-6 contact hours were more likely to have higher expectations of applying to HE than students with fewer than three hours, and contact hours greater than six did not increase the likelihood any further. Students who receive in excess of six hours of outreach may therefore accrue no further benefit on the outcome of focus (expectation of HE progression at age 18 or 19), and these incremental interactions are a potential drain on programme resource which could be better used to support engagement of learners with less than three hours of total contact.

Recommendations for HeppSY Practice

Given the findings highlighted above, two recommendations are made for future HeppSY practice. Firstly, alongside the programme aim of having over 20% of target learners engage in at least two interventions per year, these interventions should endeavour to offer at least three hours of total contact time. Secondly, school register data should be used to gradually deprioritise students who have already received over six hours of contact within the academic year, unless there is a compelling case for further specific intervention (e.g. participation in a headline programme, or diversifying the profile of activity that they have undertaken). Minimising delivery of outreach activity beyond six hours per student would reduce a source of programme 'leakage' (Harrison, 2012), where additional intervention appears to be offering minimal benefit on expectations of HE progression. However, it should be noted that the 3-6-hour window was based on exploratory analysis and should be taken as a guide rather than a precise requirement.

Recommendations for Future Research and Evaluation

The matched analysis approach used here can be extended to different outcome measures drawn from the CFE annual learner survey, such as HE knowledge and career knowledge. This cohort-level analysis could aim to evaluate the association between changes in these self-report measures and participation in specific types of activity, rather than a simple measure of total contact hours. This level of analysis may be possible where multiple year groups are included, therefore providing a larger sample size.

Additionally, access to KS2 and/or KS3 attainment data at a student level would provide an opportunity for HeppSY and other UCP consortia to substantially improve upon the strength of evidence they are able to establish. If these data become available to UCP consortia, then we would revisit our matched analysis to incorporate this important variable and also include it in future work using objective HE outcomes through the HESA track data.

References

Anders, J. and Micklewright, J., 2015. Teenagers' expectations of applying to university: how do they change?. *Education sciences*, 5(4), pp.281-305.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A. and Vignoles, A., 2013. Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(2), pp.431-457.

Gorard, S., 2014. The widespread abuse of statistics by researchers: What is the problem and what is the ethical way forward?. *Psychology of education review.*, 38(1), pp.3-10.

Harrison, N., (2012). The mismeasure of participation: how choosing the 'wrong' statistic helped seal the fate of Aimhigher. *Higher Education Review*, 45(1).

HeppSY, (2020). 2019/2020 HeppSY Partnership Report [online]. Available from: <https://www.heppsy.org/assets/2020/05/HeppSY-Partnership-Report-2019-20-final.pdf>

O'Connell, A.A., 2006. *Logistic regression models for ordinal response variables* (Vol. 146). Sage.

Appendix 1

Table 1. Model 1 Ordinal Regression Statistics.

Variable	Log odds	Std. Error	Wald
Contact Hours	0.02	0.02	1.09
HE Potential: Green ^a	0.54	0.27	4.06
HE Potential: Non-UCP ^a	0.83	0.28	8.69
Male ^b	-0.44	0.17	1.48
W1 – Definitely won't apply ^c	-1.14	0.57	3.97
W1 - Very unlikely ^c	-0.97	0.47	4.19
W1 - Fairly unlikely ^c	-0.04	0.41	0.01
W1 - Fairly likely ^c	0.70	0.28	6.11
W1 - Very likely ^c	1.62	0.29	31.14
W1 - Definitely will apply ^c	2.96	0.32	84.96

^a Reference category - HE Potential: Red

^b Reference category – Female

^c Reference category – 'Don't know' W1 responses

Table 2. Model 2 Ordinal Regression Statistics.

Variable	Log odds	Std. Error	Wald
Contact Hours	0.05	0.02	5.92
HE Potential: Green ^a	0.40	0.30	1.74
HE Potential: Non-UCP ^a	0.56	0.34	2.69
Male ^b	-0.52	0.18	8.11
W1 – Definitely won't apply ^c	-1.41	0.60	5.55
W1 - Very unlikely ^c	-1.12	0.49	5.79
W1 - Fairly unlikely ^c	0.12	0.42	0.13
W1 - Fairly likely ^c	0.78	0.29	7.00
W1 - Very likely ^c	1.65	0.30	30.08
W1 - Definitely will apply ^c	3.17	0.33	90.82

^a Reference category - HE Potential: Red

^b Reference category – Female

^c Reference category – 'Don't know' W1 responses

Table 3. Model 3 Ordinal Regression Statistics.

Variable	Log odds	Std. Error	Wald
Contact Hours	0.05	0.02	5.04
HE Potential: Green ^a	0.40	0.30	1.77
HE Potential: Non-UCP ^a	0.53	0.35	2.32
Male ^b	-0.51	0.18	7.91
W1 – Definitely won't apply ^c	-1.32	0.62	4.52
W1 - Very unlikely ^c	-1.15	0.49	5.54
W1 - Fairly unlikely ^c	0.29	0.43	0.46
W1 - Fairly likely ^c	0.74	0.30	6.21
W1 - Very likely ^c	1.60	0.30	27.65
W1 - Definitely will apply ^c	3.10	0.34	84.19
Motivation (W1)	0.12	0.11	1.22

^a Reference category - HE Potential: Red

^b Reference category – Female

^c Reference category – 'Don't know' W1 responses



To find out more please contact:

Mike Garnock-Jones

Director, Hepp

M.Garnock-Jones@shu.ac.uk

Greg Brown

Evaluation and Data Manager, HeppSY

G.Brown@shu.ac.uk

Sharon Woodward-Baker

HeppSY Programme Manager

S.Woodward-Baker@shu.ac.uk

